
GPflowOpt: A Bayesian Optimization Library using TensorFlow

Nicolas Knudde* Joachim van der Herten

Tom Dhaene Ivo Couckuyt†

Ghent University – imec

IDLab – Department of Information Technology

{nicolas.knudde, joachim.vanderherten, tom.dhaene, ivo.couckuyt}@ugent.be

Abstract

A novel Python framework for Bayesian optimization known as GPflowOpt is introduced. The package is based on the popular GPflow library for Gaussian processes, leveraging the benefits of TensorFlow including automatic differentiation, parallelization and GPU computations for Bayesian optimization. Design goals focus on a framework that is easy to extend with custom acquisition functions and models. The framework is thoroughly tested and well documented, and provides scalability. The current released version of GPflowOpt includes some standard single-objective acquisition functions, the state-of-the-art max-value entropy search, as well as a Bayesian multi-objective approach. Finally, it permits easy use of custom modeling strategies implemented in GPflow.

1 Bayesian Optimization

Bayesian Optimization (BO) is a principled way to find a global optimum of an objective function over a bounded domain, formally expressed as

$$\arg \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}). \quad (1)$$

The standard configuration for BO applies the principle of dynamic programming and sequentially generates a single candidate decision \mathbf{x}^* for evaluation. Given the expensive nature of f , the aim is to keep the number of iterations required to identify optimal values small. All previous evaluations are used to train a (Bayesian) model which supports the search for the next decision. BO frequently applies the non-parametric Bayesian models known as Gaussian Processes (GPs) (Rasmussen and Williams, 2006) to act as a surrogate of the objective function(s). To determine the next candidate an acquisition function is maximized over the compact domain (Snoek et al., 2012). This acquisition function usually maps the predictive distribution of the underlying model to a scalar value.

Several extensions have been proposed to this standard setting. Batch BO evaluates batches sequentially (Ginsbourger et al., 2010; González et al., 2016) to make use of parallel evaluation of the objective. The objective itself may be multi-dimensional for which multiple equally optimal solutions exist involving a trade-off (the Pareto front), this setting can be approached with multi-objective BO (Couckuyt et al., 2014).

*ORCID ID: 0000-0001-5322-5930

†Ivo Couckuyt is a post-doctoral research fellow of FWO-Flanders.

Single-objective	Expected Improvement (EI) (Moćkus, 1975) Lower Confidence Bound (LCB)(Srinivas et al., 2010) Max-Value Entropy Search (MES) (Wang and Jegelka, 2017) Probability of Improvement (PoI) (Kushner, 1964)
Multi-objective	Hypervolume-based PoI (HvPoI) (Couckuyt et al., 2014)
Constraint	Probability of Feasibility (PoF) (Schonlau, 1997)

Table 1: Implemented acquisition functions

2 Motivation

There are many libraries for BO available, whether it be commercial or open-source. Of the latter category Spearmint (Snoek et al., 2012) and GPyOpt (The GPyOpt authors, 2016) are well-known: both packages are written in Python and rely on NumPy for numerical operations. Alternatives include BayesOpt (Martinez-Cantin, 2014), implemented in C++, and RoBO (The RoBO authors, 2016).

These available packages are written in a modular way and are relatively easy to use, but may lack extensive documentation or rigorous testing. Furthermore, adding new acquisition functions is usually quite easy but introducing a different modeling strategy requires serious effort. Confronted with the current research challenges and novel application domains for BO this motivated the development of a new framework: we intended to create an interface that is straightforward to use, such that it becomes easy to extend the code and develop new techniques with a minimum of overhead. Finally, the recent development of computing libraries supporting automated differentiation and providing scalability provide a natural base for a Bayesian optimization package.

3 Design choices

As the choice of modeling framework has a significant impact on the design of the resulting BO framework, several alternatives were evaluated. Ultimately we chose GPflow (Matthews et al., 2017) as framework for modeling: GPs are the most common surrogate model used in BO, and GPflow makes development and implementation of custom GP models for BO considerably easier. The package is written in Python and provides a powerful framework for implementing (GP) models, including Sparse GPs and GP-Latent Variable Models, using variational inference as the standard approximate inference technique. As GPflow is built on TensorFlow, it enables use of GPU computations, parallelization and automatic differentiation.

The development resulted in the release of the open-source GPflowOpt project featuring following properties:

1. Simple application of **different models** (using the GPflow framework) as a surrogate in Bayesian Optimization,
2. **Automated differentiation** increasing ease of implementation,
3. Support for (multiple) **GPU** enabling fast computation,
4. Clean **object-oriented Python** front-end which is simple to extend,
5. Rigorous **testing** and **extensive** documentation.

The code is completely modular permitting simple implementation of different acquisition functions. Acquisition functions that are included with the framework are summarized in Table 1. Both single- and multi-objective acquisition functions are implemented as well as Probability of Feasibility (PoF) to incorporate black-box constraints. GPflowOpt supports most models included in GPflow such as the Variational Gaussian Process (VGP) or the Sparse Variational Gaussian Process (SVGP), and allows the use of custom models. The next version of GPflow intends to further increase these capabilities as integration of other TensorFlow modeling frameworks (such as Keras) with GPflow will be possible.

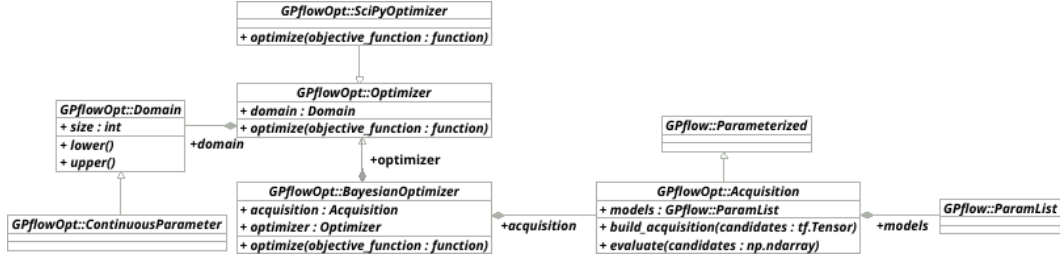


Figure 1: UML diagram of GPflowOpt

	GPflowOpt	GPyOpt	Spearmin	BayesOpt	RoBO
Language	Python	Python	Python	C++	Python
Auto differentiation	✓	✗	✗	✗	✗
Multi-objective	✓	✗	✗	✗	✗
Code coverage	99%	56%	—	—	46%
Batch BO	✗	✓	✗	✗	✗
GPU support	✓	✗	✗	✗	✗

Table 2: Comparison of Bayesian optimization frameworks. The information are taken from the respective web pages or papers (The GPyOpt authors, 2016; Snoek et al., 2012; Martinez-Cantin, 2014; The RoBO authors, 2016)

4 Implementation

Following the modular structure of GPflow, the main building blocks of GPflowOpt are Domain, Acquisition and Optimizer. The class relationships are summarized in Figure 1.

The expensive objective function f is defined by the user. The domain is a GPflowOpt class containing the bounds of the optimization domain, which is used for scaling purposes and configuration of optimizer objects. The acquisition function (Acquisition) holds one or more GPflow models and maps their predictions to a score. By default a transparent model wrapper is used for automated data scaling to and from the underlying model to increase the success rate of the hyperparameter optimization. The BayesianOptimizer class handles the classic BO process and includes model optimization, optimization of the acquisition function, evaluation of the objective and optionally marginalization of hyperparameters using Hamiltonian Monte-Carlo sampling.

5 Comparison

In Table 2, a comparison is made to the popular BO frameworks summarized in Section 2. Most frameworks are written in Python. Key features of GPflowOpt over other frameworks include the support for multi-objective objective functions along with an implementation of an acquisition function specifically for this type of applications, the rigorous testing suite resulting in a code coverage of 99% and extensive documentation. Additionally, a fast algorithm for generating maximin Latin hypercubes Viana et al. (2010) is included. Latin hypercubes are a popular method for generating space-filling Design of Experiments (DoE) to start the Bayesian optimization process.

On the other hand GPyOpt supports batch BO, which is currently still in development for GPflowOpt and was not part of the first release. In terms of efficient hardware usage a native implementation in C++ as offered by BayesOpt is preferable, however the computations in GPflowOpt are carried out by TensorFlow graphs which benefit of a native computational back-end. The additional scalability and automated differentiation compensate for the overhead of the framework. Another powerful feature of GPflowOpt is the option to implement models in the GPflow framework, allowing their use without the need of implementing wrapper classes due to the modular structure.

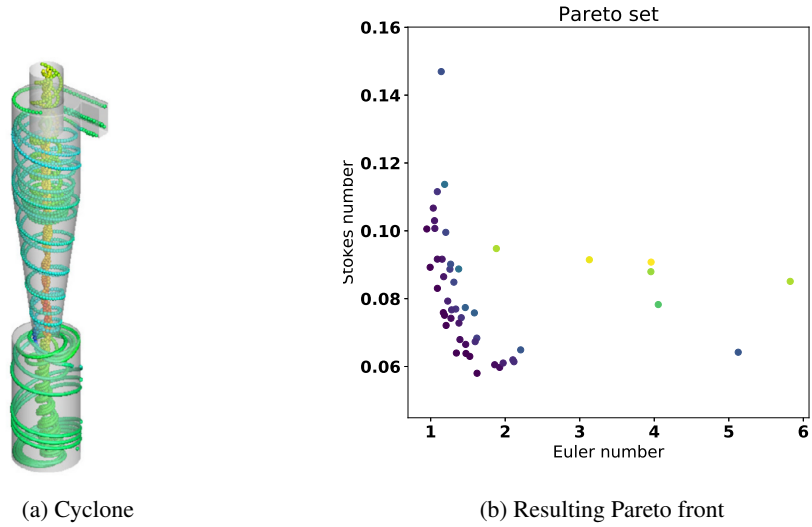


Figure 2: Multi-objective optimization of a cyclone separator (shown on the left). The Pareto front only depicts the feasible samples in the objective space. The bottom-left corner represents the optimal position, and the colors indicate its dominance.

6 Illustration

The following example presents the optimization of a gas cyclone separator (depicted in Figure 2a), a real-world device that is able to separate dust particles from gases through a complex swirling motion. Finding an optimal solution involves a trade-off between two conflicting objectives which will be solved with multi-objective BO. The device is characterized by 7 geometric parameters, which are the input of the expensive objective function which results in two objectives: the pressure loss (represented by the Euler number) and the cut-off diameter (represented by the Stokes number). The multi-objective BO approach used here is the Hypervolume Probability of Improvement (HvPoI) (Couckuyt et al., 2014), which indicates the probability of a candidate evaluation improving the volume between the Pareto front and a reference point (e.g., the anti-ideal point) in the objective space. At the same time four production inequality constraints based on the same inputs have to be taken into account. These constraints are black-box themselves and are modeled as well. By incorporating the Probability of Feasibility (PoF) (Schonlau, 1997) into a joint acquisition function this aspect can be included, as the feasibility is learnt jointly with the objectives.

A maximin Latin hypercube consisting of 50 points was used as an initial design and is implemented by using the Translational Propagation algorithm (Viana et al., 2010). A total of 120 evaluations of the objective function were performed, each evaluation yields both the Euler and Stokes numbers, as well as the constraint values. The resulting Pareto front of the feasible samples is shown in Figure 2b.

7 Conclusion and future work

A new, versatile Python package for Bayesian optimization was introduced. It allows for straightforward TensorFlow model integration by building on GPflow. This offers significant benefits such as automatic differentiation, multi-core calculations and GPU support. A comparison was made with other open-source packages indicating the package offers significant advantages including rigorous testing and extensive documentation. Currently, the framework still lacks advanced sampling techniques such as batch BO. The current short-term roadmap of GPflowOpt is batch BO as well as support for discrete and categorical variables.

The development of GPflowOpt is open-source and fully transparent. Hence, the scientific community is encouraged to make contributions to the framework and test out their own Bayesian optimization algorithms.

References

- Couckuyt, I., Deschrijver, D., and Dhaene, T. (2014). Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. *Journal of Global Optimization*, 60(3):575–594.
- Ginsbourger, D., Le Riche, R., and Carraro, L. (2010). Kriging is well-suited to parallelize optimization. *Computational Intelligence in Expensive Optimization Problems*, 2:131–162.
- González, J., Dai, Z., Hennig, P., and Lawrence, N. (2016). Batch bayesian optimization via local penalization. In *Artificial Intelligence and Statistics*, pages 648–657.
- Kushner, H. J. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 86(1):97–106.
- Martinez-Cantin, R. (2014). BayesOpt: A Bayesian Optimization Library for Nonlinear Optimization, Experimental Design and Bandits. *Journal of Machine Learning Research*, 15:3915–3919.
- Matthews, A. G. d. G., van der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., León-Villagrà, P., Ghahramani, Z., and Hensman, J. (2017). GPflow: A Gaussian Process Library using TensorFlow. *Journal of Machine Learning Research*, 18(40):1–6.
- Moćkus, J. (1975). On Bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference*, pages 400–404. Springer.
- Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian Processes for Machine Learning*, volume 1. MIT press Cambridge.
- Schonlau, M. (1997). Computer experiments and global optimization.
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959.
- Srinivas, N., Krause, A., Seeger, M., and Kakade, S. M. (2010). Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In Fürnkranz, J. and Joachims, T., editors, *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 1015–1022. Omnipress.
- The GPyOpt authors (2016). GPyOpt: A Bayesian Optimization framework in python. <http://github.com/SheffieldML/GPyOpt>.
- The RoBO authors (2016). RoBO: a Robust Bayesian Optimization framework. <http://github.com/automl/RoBO>.
- Viana, F. A., Venter, G., and Balabanov, V. (2010). An algorithm for fast optimal Latin hypercube design of experiments. *International Journal for Numerical Methods in Engineering*, 82(2):135–156.
- Wang, Z. and Jegelka, S. (2017). Max-value Entropy Search for Efficient Bayesian Optimization. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3627–3635, International Convention Centre, Sydney, Australia. PMLR.